

Accurate Garment Surface Analysis using an Active Stereo Robot Head with Application to Dual-Arm Flattening

Li Sun¹, Gerardo Aragon-Camarasa¹, Simon Rogers¹, J. Paul Siebert¹

Abstract— We present a visually guided, dual-arm, industrial robot system that is capable of autonomously flattening garments by means of a novel visual perception pipeline that fully interprets high-quality RGB-D images of the clothing scene based on an active stereo robot head. A segmented clothing range map is B-Spline smoothed prior to being parsed by means of shape and topology into ‘wrinkle’ structures. The wrinkle length, width and height are used to quantify the topology of wrinkles and thereby rank the size of wrinkles such that a greedy algorithm can identify the largest wrinkle present. A flattening plan optimised for this specific wrinkle is formulated based on dual-arm manipulation. Validation of the reported autonomous flattening behaviour has been undertaken and has demonstrated that dual-arm flattening requires significantly fewer manipulation iterations than single-arm flattening. The experimental results also reveal that the flattening process is heavily influenced by the quality of the RGB-D sensor, use of a custom off-the-shelf high-resolution stereo-based sensor system outperforming a commercial low-resolution kinect-like camera in terms of required flattening iterations.

I. INTRODUCTION

Autonomous manipulation of garments still remains an extremely difficult task for autonomous robotic systems. This is due to the perceptual and manipulation abilities required while interacting with these highly deformable objects. In this paper, we intend to advance the state-of-the-art in perception and manipulation of garments by improving: perceptual robotic competence by means of an actuated binocular robot head system (Fig. 1-B), the description and quantification of the garment surfaces shapes and topologies, and the manipulation skills of a dual-arm robot (Fig. 1-A).

In this paper, we describe an active stereo sensing system for automatic gaze control, camera vergence and GPU accelerated stereo matching in order to support a 16MP high quality point cloud stream at 0.2Hz. We also present a novel and generic 2.5D deformable surface analysis approach for parsing, characterising and quantifying a garment’s surface. To demonstrate and validate our surface analysis approach and our active stereo sensing system, we propose an on-table garment flattening robotic task. For this purpose, we have devised a dual-arm flattening strategy in order to flatten a variety of garments in real-life scenarios. Videos of our experimental results are available at:

<http://youtu.be/Z85bW6QqdMI>

II. RELATED WORK

Maitin et al. [1] developed one of the first successful robotic systems to manipulate garments - specifically towels.

^{*}European FP7 Strategic Research Project, CloPeMa; www.clopema.eu

¹School of Computing Science, University of Glasgow, 17 Lilybank Gardens, G12 8RZ, Glasgow, UK l.sun.1@research.gla.ac.uk

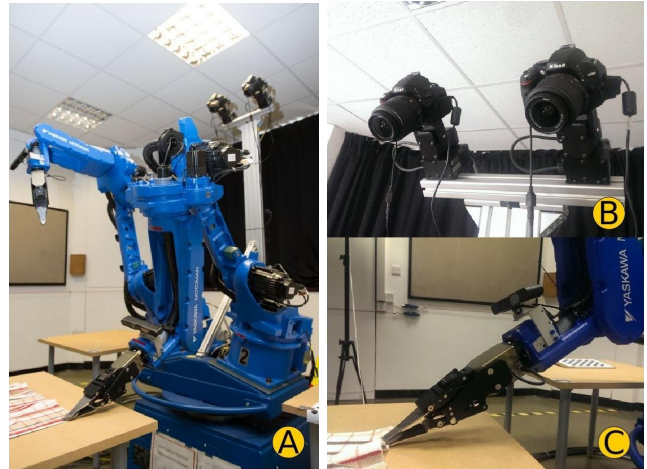


Fig. 1. (A) The CloPeMa robot which consists of two seven degrees of freedom YASKAWA arms and a custom made YASKAWA turn-table. Each arm features a specialised gripper for handling clothing and a ASUS Xtion Pro. (B) Our stereo robot head instagrated on the CloPeMa testbed suit. (C) A close up of the CloPeMa gripper.

They use a combination of stereo cameras and depth sensors in order to search for geometric cues that facilitate the detection of corner grasp points on towels, an in-air unfolding strategy, and towel folding and piling. Since then, progress over different clothing scenarios has been made. These include: optimal garment grasping [2], [3], garment recognition [4], [5], pose estimation [6], [7] garment separation [8], garment folding [9]–[11] and unfolding [12]–[15].

Common ground in the cited research lies in the selection of the imaging system and the perceptual capabilities of the robot. Imaging systems comprise kinect-like sensors [2], [3], [8], [11], [13]–[16] and poor-quality stereo systems for near real-time operation [4], [5], [13]. However, these type of sensors limit the perceptual competence of the robot due to the fact that depth sensors provide low-resolution depth maps and suffer from image noise (e.g. Fig. 3). These limitations make characterising the surface topology and state of a garment with high accuracy challenging. Similarly, current research has focused on visual perception approaches that are limited in scope and constrained to the task at hand. Therefore, existing visual perception approaches for clothing are in short of generic surface shape and topology analysis. For example, researchers have described the garment in terms of polygonal representations [9]–[11], [16] which are constrained to a specific task.

For the specific garment flattening robotic task tackled in this paper, current research falls into two categories: *gravity-*

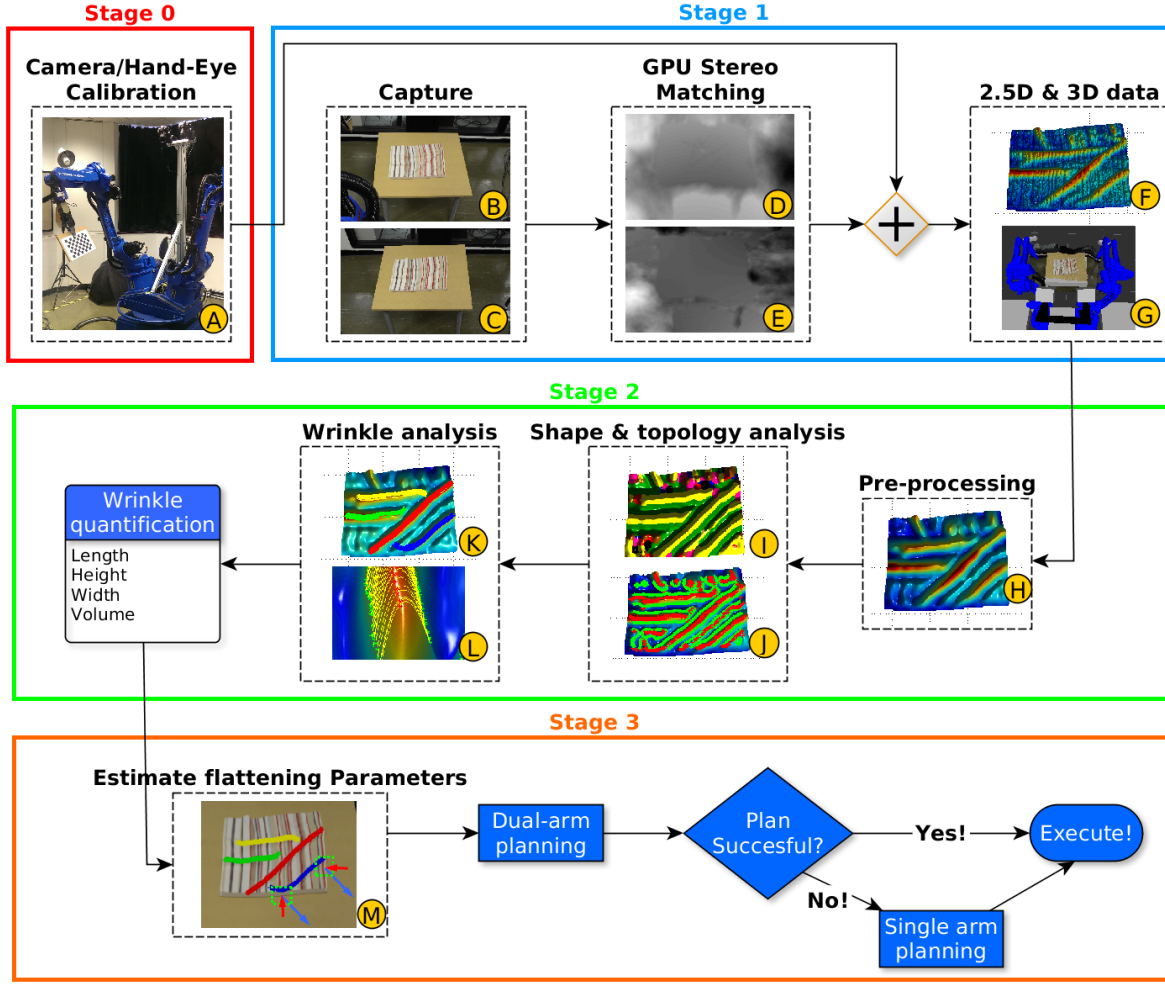


Fig. 2. Overall processing pipeline for garment perception and manipulation.

based flattening [14], [15] or by sliding the garment along a table [12]. In both cases, the garment cannot be guaranteed to be flattened; this is because these approaches [9]–[12] do not track the state of the garment in accordance to the perception-action loop principle. Overall, we conclude that the current state-of-the-art in garment manipulation requires high-resolution imaging systems for garment perception and, therefore, the understanding of the garment state for competent actions. To the authors’ knowledge, this paper is the first study that successfully proposes a generic 2.5D description and quantification method that *fully* characterises the garment’s high-dimensional configuration space. Even though our generic approach can be used for other types of manipulation in garments (e.g. optimal grasp point selection), we have selected the garment flattening problem because this task requires understanding the garment’s state for each manipulation. This task also allows us to demonstrate our stereo-head, generic surface analysis and the dual-arm flattening strategy working together under the perception-action loop.

III. METHODOLOGY

As observed in Figure 2, our processing pipeline consists of four stages: (0) camera calibration and integration (of-line), (1) depth data acquisition, (2) generic garment surface analysis, and (3) dual-arms planning for flattening. The following subsections detail the design and implementation of each stage. We assume that the garment has been selected, separated and unfolded from a pile of clothes and has been laid on a table. It must be noted that this pipeline has been implemented in the Robot Operating System (ROS Hydro) [17] and it is freely distributed as per the CloPeMa project license agreement¹.

A. Active Binocular Robot Head

To allow our dual-arm robot to parse and fully describe the surface topology of garments, we designed an actuated binocular robot head. The robot head comprises two off-the-shelf Nikon DSLR cameras (D5100) that capture images at a resolution of 16 mega pixels (MP). Each camera is mounted

¹<http://www.clopema.eu>

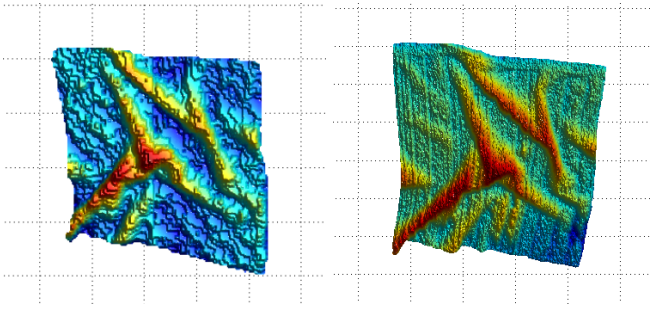


Fig. 3. An example of a 2.5D range map as captured by (left) a RGBD sensor - ASUS Xtion Pro - and (right) by our stereo head. Robot head data provides better quality and larger resolution than the depth sensor. For more examples, please visit our stereo-matching benchmark dataset: <https://sites.google.com/site/ugstereodatabase/> [22]

on two degrees of freedom pan and tilt units (PTU-D46) which are separated by a pre-defined baseline for stereo capturing. The hardware is interfaced to an Intel Core i7-3930K computer at 3.20 GHz with 32GB of RAM running Ubuntu 12.04 and ROS. Fig. 1-B depicts the robot head as mounted on our dual-arm robot. The robot head's visual capabilities include: camera vergence, automatic gaze control and visual search based on SIFT features [18], multiple same-class instance recognition [19] and GPU stereo matching and point cloud/depth map computation [20] (with a processing rate of 0.2 fps on 16MP images).

In order to maintain calibration of the camera system under extrinsic camera movement in our binocular robot head, we interfaced OpenCV calibration routines² to find the intrinsic camera parameters of each camera. We also integrated Tsai's hand-eye calibration routines to estimate rigid geometric transformations between PTU to camera and gripper to camera - stage 0 in Fig. 2. Knowledge of the intrinsic and extrinsic camera parameters and geometric transformations of cameras, PTUs and the robot allow us to update the stereo relationship of the camera system and, therefore, complete the robot's kinematic structure. It is thus possible to direct its gaze under program control and recover metrically accurate depth images under dynamic movement of the camera system and, therefore, map 2.5D information into the robot coordinate frame Fig. 2-F and G. Depth map and point cloud computations are supported on the stereo head by means of a GPU accelerated stereo matcher [20] and a simplified least-square stereo triangulation routine [21]. Fig 3 shows a visual comparison of depth maps between our robot head and the ASUS Xtion.

B. Pre-processing

Before analysing the garment's surface topology, it is necessary to separate the imaged garment from the background (pre-processing step - Fig. 3-H). For this purpose, we adopt and modify a previously reported grab-cut segmentation approach based on Gaussian mixture models [11]. Image segmentation is thus crafted to the task on-hand, namely,

garment flattening. After segmentation, the robot's field of view consists of three different regions: the garment, the operating table and the background. We therefore focus the attention of the robot on the garment for the analysis and quantification of the garment's surface topology.

As geometry based depth features (e.g. curvature and shape index) are extremely susceptible to high frequency noise, a piece-wise B-Spline surface approximation is used to fit a continuous implicit surface onto the original depth map (an example is shown in Fig. 2-H). In our approach, we first divide the depth map of the current view into square patches according to the x-y depth-map plane. We then fit an open uniform B-Spline surface to each patch [23]. Adjacent patches are connected together by blending the B-Spline's control points in order to ensure continuity³.

C. The 2.5D deformable surface analysis approach

Our surface analysis approach consists of: *garment shape and topology analysis*, and *wrinkle description and quantification*. Both are described in the following subsections.

1) *Garment Shape and Topology Analysis*: Surface topographic features are detected by computing the curvatures of the garment surface. For this purpose, we employ *shape index* [23] to classify the surface shape at each point into one of nine different types: *cup*, *trough*, *rut*, *saddle rut*, *saddle*, *saddle ridge*, *ridge*, *dome* and *cap* (as shown in Fig. 2-I). A majority ranking filtering is applied to remove incorrect and isolated shapes. Wrinkles are finally parametrised in order to be used in the manipulation task. Our shape analysis approach focuses on wrinkles (yellow regions in Fig. 2-I) because wrinkles determine the configuration of a garment irrespectively of its colour, fabric or garment type. Others shapes such as *dome* and *rut*, representing the shapes of wrinkles at junctions, are adapted to separate the cross-like wrinkles in highly wrinkled situations (as shown in Fig. 4).

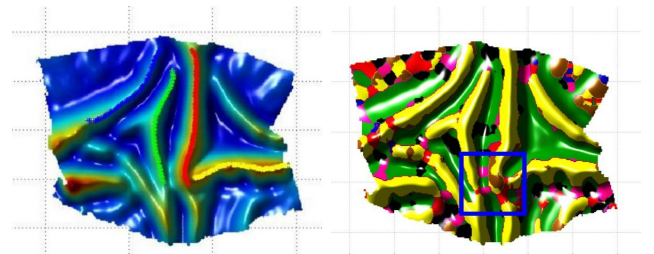


Fig. 4. In highly wrinkled situations, the shape of wrinkles at junctions are classified as *dome* or *rut* (as shown in brown and red colours); this classification is used to separate jointed wrinkles in our approach.

³The patch connecting process comprises three steps: The first step achieves C0 continuity on the patch boundaries by sharing control points. Then, the cross boundary derivatives (twist vector) are enforced to be C1 continuous by averaging the twist vectors. The boundary control points are finally made C1 continuous along the vertical and horizontal directions. In our implementation, a 3rd order uniform open knot vector [0 0 0 0 1 2 2 2 2] is used to compute the basis function. Each patch is controlled by 5×5 control points in order to achieve C2 continuity. At the connecting step, control points are adjusted on the boundaries to achieve C1 continuity between adjacent patches. Hence, C1 continuity is ensured after B-Spline surface fitting.

²<http://opencv.org>

In order to detect and locate wrinkles on the garment's surface, we first provide a geometric definition of a wrinkle (Definition 1). This definition is inspired by classical geometric approaches for parsing 2.5D surface shapes and topologies (e.g. shape index [23]). As a consequence of Definition 1, it is essential to define what ridges are as these encapsulate most of the wrinkle's topographic information (Definition 2).

- **Definition 1.** A Wrinkle is a set of triplets. Each triplet consists of one ridge point and two contour points along the maximal curvature direction.
- **Definition 2.** Ridge points are the positive extrema of maximal curvature [24]. The wrinkle's contour is the boundary of the concave and convex surfaces of the garment.

In this paper, the definition of *ridges* shares similarities to that given by Ohtake et al. [25], the only difference is that instead of estimating curvatures from a polygon mesh, we calculate these surface curvatures using differential geometry, obtained directly from the depth map⁴.

The wrinkle's contour is defined as the boundary of the convex and concave surface (as per Definition 2), which is estimated by computing the zero-crossing of the second derivatives of the garment's surface⁵. After the wrinkle's contour has been detected, the garment surface topology is fully parsed, an example of which can be seen in in Fig. 2-I and J.

2) *Wrinkle Description:* The wrinkle description process consists of: (1) connecting ridge points to contiguous segments and (2) grouping found segments into wrinkles (Fig. 2-H). In the first step, neighbouring pixels of detected ridges are connected iteratively until continuous segments are obtained. In this process, the segments' end points are labeled as 'active' ridge points. The minimum l_2 distance between every two active ridge points is used as the distance measurement between two segments (Eq.1):

$$d(s_i, s_j) = \begin{cases} \min \| \forall r_m^a \in s_i, \forall r_n^a \in s_j \|_2, & \text{if } r_m, r_n \in \gamma \\ \infty, & \text{otherwise} \end{cases}, \quad (1)$$

where s_i and s_j are two ridge line segments consisting of a range of active ridge points $\{r_1^a, \dots, r_M^a\}$ and $\{r_1^a, \dots, r_N^a\}$, and γ is a *ridge* region in the shape index map. If two nearest segments are in the same *ridge* region, they are grouped into a larger segment. Finally, for each wrinkle, a 5th polynomial curve is fitted along its ridge points, and this polynomial curve is the final description of a wrinkle.

⁴The details of computing curvatures from depth is described as follows. The 2.5D points in the depth map are examined pixel by pixel in order to find if they are the positive extrema along the maximal curvature direction. That is, given a depth map I , for each point p in I , the mean curvature C_m^p and Gaussian curvature C_g^p are firstly calculated as: $C_m^p = \frac{(1+(f_x^p)^2)f_{xx}^p + (1+(f_y^p)^2)f_{yy}^p - 2f_x^p f_y^p f_{xy}^p}{2(\sqrt{1+(f_x^p)^2 + (f_y^p)^2})^3}$ and $C_g^p = \frac{f_{xx}^p f_{yy}^p - (f_{xy}^p)^2}{(1+(f_x^p)^2 + (f_y^p)^2)^2}$,

where first derivatives f_x^p, f_y^p , and second derivatives $f_{xx}^p, f_{yy}^p, f_{xy}^p$ are computed. Then, the maximal curvature can be calculated by C_m^p and C_g^p as: $k_{max}^p, k_{min}^p = C_m^p \pm \sqrt{(C_m^p)^2 - C_g^p}$.

⁵In our implementation, we apply a Laplacian template window size of 16×16 in order to calculate the second order derivative.

3) *Wrinkle Quantification:* Shape Index classifies surface shapes without measuring surface magnitude. We therefore measure a wrinkle's surface magnitude in terms of triplets, as stated in Definition 1. By matching ridge points and their two corresponding contour points (as shown in Fig. 2-L), we estimate the maximal curvature direction θ (on the x-y plane of the depth map) on each ridge point. The maximal curvature direction, θ , is estimated with respect to the perpendicular direction of the tangent line on the fitted wrinkle. In our implementation, we adopt a 5th order polynomial curve in order to approximate each wrinkle⁶. Given a ridge point p_r in the depth map I , we search over the two directions defined by θ and its inverse direction until the corresponding wrinkle's contour points p_c^l and p_c^r are found. If the surveyed patch is in the same *ridge* region as p_r (shown as yellow in Fig.2-K), the search continues, otherwise the search terminates.

In our approach, the three points that define a triplet are used to measure the height and width of a wrinkle. That is, given a triplet t_p containing one ridge point p_r and two contour points p_c^l and p_c^r , the height h_t and width w_t are calculated as follows:

$$h_t = 2 \frac{d(d-a)(d-b)(d-c)}{c} \quad (2)$$

$$w_t = c, \quad (3)$$

where $a = \|p_r, p_c^l\|_2$, $b = \|p_r, p_c^r\|_2$, $c = \|p_c^l, p_c^r\|_2$, and $d = (a + b + c)/2$. The numerator of Eq. 2 is the area of a triangle embedded in a 3D space.

For the specific task of garment flattening, the physical volume of the wrinkle is adopted as the score for ranking detected wrinkles. We also apply PCA on the largest wrinkle for the current view in order to infer the two grasping points and the flattening directions for each arm. To obtain the amount upon which the dual-arm robot should pull in order to remove the selected wrinkle, we compute the geodesic distance between the two contour points of each triplet [26]⁷. The following section details how we used these estimated parameters for flattening a garment.

D. Dual-Arm Flattening

1) *Flattening Heuristic:* For the heuristic flattening strategy, we adopt a greedy search approach, in which the largest wrinkle detected is eliminated in each perception-action iteration. **We consider the largest wrinkle detected per flattening iteration because the entropy added into the system increases when considering a group of wrinkles with similar directions and the likelihood that dual-arm planning would succeed is significantly reduced. Therefore, the largest wrinkle detection heuristic guarantees**

⁶The curve function is defined as; $f(x) = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$, and, the tangent direction, $\alpha = \arctan(5ax^4 + 4bx^3 + 3cx^2 + 2dx + e)$.

⁷We used Gabriel Peyre's toolbox for the implementation of the geodesic distance: <http://www.mathworks.co.uk/matlabcentral/fileexchange/6110-toolbox-fast-marching>

that a solution is achieved regardless highly wrinkled situations. Wrinkles are quantified according to their physical volume. The volume of a wrinkle w is thus given by:

$$volume_w = \log(l_r) + \log((\sum_{t_i \in w}^{N_r} w_t \times h_t)/N_t), \quad (4)$$

where N_r is the number of fitted ridge points in w , N_t refers to the number of matched triplets, t_i is the i th triplet of w , w_t and h_t refers to the width and height of the triplet t_i , and l_r is the length of the wrinkle which is calculated by summing up the l^2 distance between every two nearest ridge points.

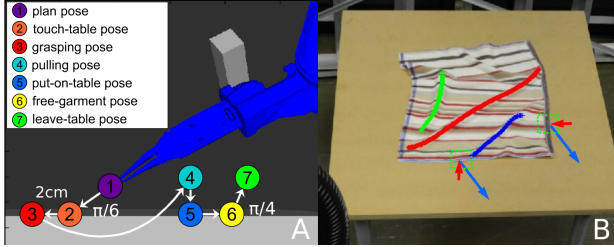


Fig. 5. (A) The 7 poses for a robotic flattening motion. The sequence of these poses is as follows: The gripper is moved to the ‘plan pose’, from this pose, we interpolate among poses sequentially in order to move the gripper(s). It is noticeable that the grasping direction and pulling direction are not aligned, the ‘plan pose’, ‘touch-table pose’, ‘grasping pose’ are coplanar, while the ‘grasping pose’, ‘pulling pose’, ‘put-on-table pose’, ‘free-garment pose’ and ‘leave-table pose’ are coplanar. For the gripper state, it will be set ‘open’ in ‘plan pose’, ‘close’ after ‘grasping pose’ and ‘open’ again after ‘put-on-table’ pose. (B), the three largest wrinkles are shown, where the red one is the largest. The inferred grasping and flattening (pulling) directions are shown as red and blue arrows respectively.

2) *Dual-arm flattening*: An entire flattening action consists of 7 robotic poses: *plan*, *touching-table*, *grasping*, *pulling*, *put-on-table*, *free-clothing*, *leave-table*. These poses are illustrated in Fig. 5-A. This figure also includes other pre-defined parameters used during the flattening task, e.g. orientation of the gripper with respect to the table. The *starting pose* (Fig. 5-A, purple) refers to moving the robot’s gripper close to the table in preparation for flattening, the *touching table* and *grasping poses* (Fig. 5-A, orange and red respectively) comprises grasping the garment’s boundary by interpolating the robot’s motion between these two poses. The *pulling* and *put-on-table poses* pull the grasped garment according to the geodesic distance (Eq. 5) and smoothly return the garment on the table. Finally, the *free-garment* and *leave-table poses* are for freeing the garment and leaving the table.

In order to calculate and interpolate these robotic poses, four parameters are required - *grasping position*, *grasping direction*, *flattening direction* and *flattening distance*. The ‘grasping’ and ‘pulling poses’ are estimated using these parameters. Then the other poses are inferred from the ‘grasping’ and ‘pulling poses’. By interpolating these 7 poses sequentially, the robot is therefore able to perform a smooth flattening action.

3) *Setting Flattening Parameters*: Here we present how to set the four mentioned flattening parameters as described in Section III-D.2. As shown in the Fig. 5-B, once the largest wrinkle is selected, PCA is employed to compute its primary direction, and the two ‘flattening directions’ are orthogonal with respect to the primary direction. After the ‘flattening directions’ are fixed, the two corresponding cross points on the garment contour are set as the position of the ‘grasping pose’ (Fig. 5-A). While the ‘grasping direction’ is estimated by the local contours of the grasping positions (as shown in Fig. 5-B). The ‘flattening distance’ d_{w_i} of wrinkle w_i is estimated by:

$$d_{w_i} = \sum_{t_i \in w}^{N_r} (G(c_l^{t_i}, c_r^{t_i}) - E(c_l^{t_i}, c_r^{t_i})) / N_t * Coeff_{spring}, \quad (5)$$

where t_i is the i th N_r triplets in w_i ; $c_l^{t_i}$ and $c_r^{t_i}$ are its two wrinkle contour points; G refers to Geodesic distance, while E refers to Euclidean distance. $Coeff_{spring}$ is the maximal distance constraints between particles in a mass-spring cloth model⁸.

Because of the limitation of the robot’s joints and possible collisions between the two arms, not all of the interpolated motions can be planned successfully. We therefore proposed a *greedy pose/motion exploration strategy* (Algorithm 1). This results in a significant improvement while flattening with both arms. However, if this algorithm fails, the robot only employs one arm - the arm used is selected according to the flattening direction.

IV. EXPERIMENTS

A. Bench-Mark Flattening Task

To validate our active stereo robot head, the 2.5D deformable surface analysis approach and the dual-arm flattening strategy, we propose two sets of experiments. Firstly, we perform 8 benchmark experiments to verify the performance and reliability while flattening a single wrinkle using dual-arm planning (Section IV-A). While, in Section IV-C, we demonstrate the performance of our proposed approach while flattening a highly wrinkled garment while comparing our robot stereo head system and standard depth cameras (i.e. kinect-like cameras). The ultimate aim of these experiments is to evaluate the performance of the methods described in this paper under pre-defined wrinkles set at different directions.

As described in previous sections, our proposed generic garment surface analysis is able to detect wrinkles that are barely discerned by human eyes unless close inspection on the garment is carried out. As it is not necessary to flatten these type of wrinkles, we therefore propose a halting criterion that determines the amount of ‘flatness’ based on the amount of the pulling distance obtained in Eq. 5. In our experiments, if detected wrinkles are less than 0.5CM (barely perceptible), the garment is considered to be flattened⁹.

⁸From our practical experience, we set $Coeff_{spring}$ as 1.10 in our experiments.

⁹This value is obtained by the average of manually flattened garment examples performed by a human-user.

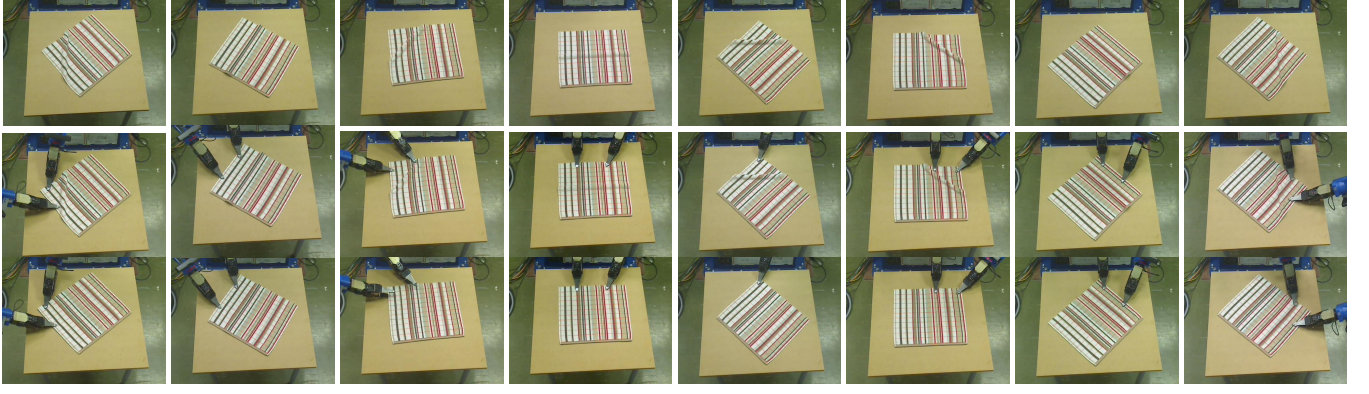


Fig. 6. Each column depicts an experiment, in which the top images show the stage before flattening, middle, flattening and bottom, after flattening.

TABLE I
THE REQUIRED NUMBERS OF ITERATIONS IN THE EXPERIMENTS.

| Bench-Mark Experiments | exp1 | exp2 | exp3 | exp4 | exp5 | exp6 | exp7 | exp8 | average |
|------------------------|------|------|------|------|------|------|------|------|---------|
| RNI | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Dual-arm Success Rate | 100% | 100% | 80% | 100% | 0% | 100% | 100% | 100% | 85% |
| Grasping Success Rate | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

TABLE II
THE REQUIRED NUMBERS OF ITERATIONS (RNI) FOR FLATTENING HIGHLY WRINKLED EXPERIMENTS. SEE TEXT FOR A DETAILED DESCRIPTION.

| Flattening Wrinkled Cloth Experiments | exp1 | exp2 | exp3 | exp4 | exp5 | exp6 | exp7 | exp8 | exp9 | exp10 | AVE | STD | Dual-Arm Success Rate |
|---------------------------------------|------|------|------|-------|------|-------|-------|-------|------|-------|-----------------|-------------------|-----------------------|
| RNI of Dual-Arm (RH) | 4(4) | 5(4) | 6(4) | 5(4) | 4(3) | 5(3) | 4(2) | 5(2) | 3(2) | 6(3) | 4.7(3.1) | 0.95(0.88) | 65.9% |
| RNI of Dual-Arm (Xtion) | 7(4) | 8(4) | 7(3) | 12(4) | 8(4) | 13(7) | 11(3) | 10(5) | 9(5) | 10(5) | 9.5(4.4) | 2.07(1.17) | 46.3% |
| RNI of Single-Arm (RH) | 7 | 12 | 5 | 8 | 7 | 7 | 12 | 14 | 8 | 6 | 8.6 | 2.99 | - |
| RNI of Single-Arm (Xtion) | 10 | 12 | 17 | 11 | 12 | 19 | 13 | 12 | 11 | 14 | 13.1 | 2.85 | - |

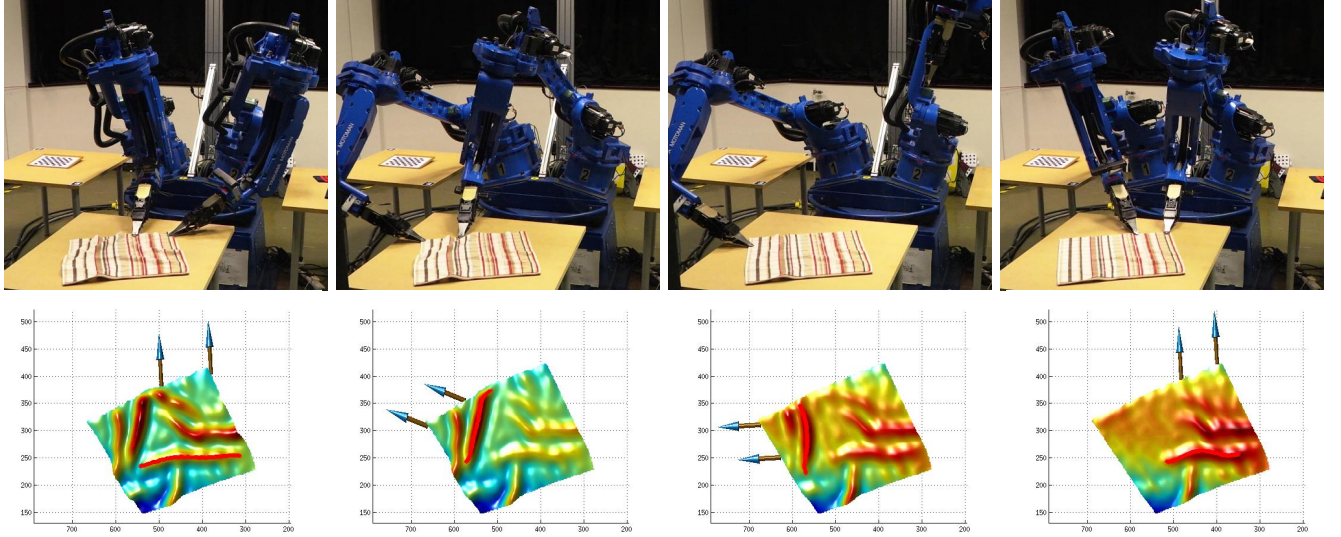


Fig. 7. The first row depicts the towel state before flattening; second row, the detected largest wrinkles and the inferred forces. On the third iteration, dual-arm planing proves infeasible to execute, a single-arm manoeuvre is then formulated and applied.

B. Benchmark experiments

As shown in Fig. 6, there is one salient wrinkle distributed in the range of 45 degree to -45 degree (from the robot's view). In order to evaluate the stability of the proposed approach, we repeat each experiment 5 times and results are

shown in Table I.

From Table I, we deduce that our generic garment surface analysis is able to flatten these 8 bench-mark experiments in only 1 iteration. Moreover, the successful rate for dual-arm planing is 85% while the robot grasps successfully the edge of the garment. Experiment 5 shows a fail case while using

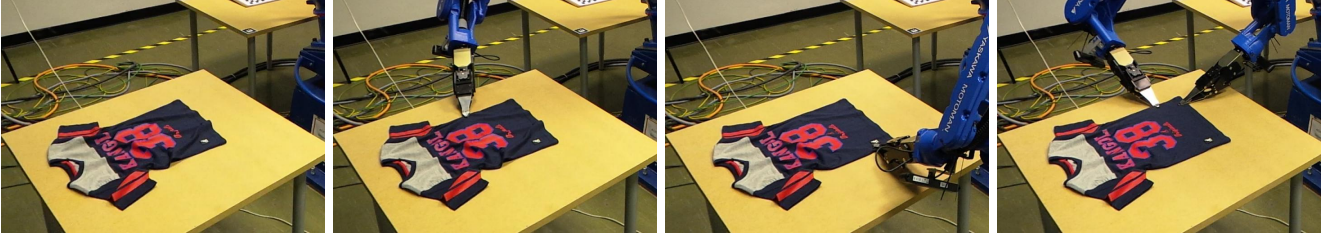


Fig. 8. Preliminary results while flattening a T-shirt. As observed, our proposed pipeline is able to adapt to any shape of garment, the robot can grasp the sleeves and stretch the wrinkles successfully.

Algorithm 1 The Pose Exploration Algorithm for Planing Dual-Arms Grasping.

In: The direction interval is d_I . The maximum numbers of exploration in each side N_E .

Out: The final planable grasping directions of two arms d_L, d_R .

Compute the ideal grasping directions D_L, D_R .

if D_L, D_R is planable **then**

$d_L = D_L, d_R = D_R$;

return d_L, d_R

end if

Set the minimal whole error of two arms $e_{min} = \infty$

for $d_l = 0$; $d_l \leq d_I \times N_E$; $d_l = d_l + d_I$ **do**

for $d_r = 0$; $d_r \leq d_I \times N_E$; $d_r = d_r + d_I$ **do**

Compute the error of left arm and right arm, $e_l = d_l/d_I, e_r = d_r/d_I$;

Compute the whole error $e_{lr} = e_l \times e_r$;

if d_l, d_r is planable and $e_{lr} < e_{min}$ **then**

$d_L = d_l; d_R = d_r; e_{min} = e_{lr}$;

end if

end for

end for

return d_L, d_R

both arms, this is due to the limitation of the robot joints and the inverse kinematic planner adopted.

C. Highly-Wrinkled Garment Flattening

In order to investigate the overall performance of our proposed approach for autonomous flattening of highly wrinkled garments, we compared the flattening performance between a single-arm and dual-arm strategies. Similarly, in order to demonstrate the use of high-quality sensing capabilities during the perception-action loop, we compare the flattening performance between our stereo robot head and a kinect-like sensor (in this paper we used the Xtion sensor¹⁰).

Therefore, for each experiment, we randomly wrinkled a squared towel - wrinkles are distributed in different directions without following an order. We then apply different flattening strategies (dual-arm or single arm) with either the robot stereo head or Xtion. For comparison, 4 groups

of experiments are carried out: (1) dual-arm using robot head, (2) single-arm using robot head, (3) dual-arm using Xtion and (4) single-arm using Xtion. To measure the overall performance and reliability, we perform 10 experiments for each group and we compute the Required Number of Iterations (RNI) as shown in Table II. In Table II, each column represents the experiment number for each of the groups proposed above. Values in parentheses depict the number of iterations where dual-arm planning was successful while the rest of the values depict the RNI for each experiment.

From Table II, we can observe that the average RNI for dual-arm flattening using robot head is 4.7 (achieving 65.9% arm planning success rate) while single-arm is 8.6. Whereas, dual-arm flattening using Xtion, the robot required 9.5 iterations (achieving 46.3% dual-arm planning success rate), while single-arm, 13.1 iterations. We can then conclude that a dual-arm strategy achieves a much more efficient performance on flattening with respect to a single-arm strategy. The standard deviation (STD) of each group of experiments are also calculated, where the STD for dual-arm flattening are 0.95 (using robot head) and 2.07 (using Xtion) while for single-arm are 2.99 and 2.85, respectively. As expected, dual-arm strategy is not only more efficient but also more stable than single-arm strategy. On the other hand, from the sensors' perspective, we find that while using the stereo robot head, the robot is able to complete successfully a flattening task in 4.7 iterations (dual-arm case) as opposed to 9.5 iteration while using Xtion. Overall, our robot head clearly outperforms the Xtion in both dual-arm flattening and single-arm flattening experiments.

From the results described above, we discovered that the dual-arm strategy is more efficient in flattening long wrinkles than the single-arm because it usually breaks long wrinkles into two short wrinkles. Likewise, between the stereo robot head and Xtion, we find that it is difficult to quantify the wrinkles and also estimate the accurate flattening displacement (especially for small wrinkles) from pure Xtion depth data because the depth map is noisier than the robot head - the high frequency noise is usually more than 0.5 CM. Furthermore, long wrinkles captured by the Xtion are often split into two small wrinkles due to the poor quality of the depth map, which in turn results in more flattening iterations (and lower dual-arm planing success rate). A video example of the above experiments can be found at: <http://youtu.be/Z85bW6QqDMI>

¹⁰http://www.asus.com/uk/Multimedia/Xtion_PRO_LIVE/

V. CONCLUSIONS

In this paper, we described an active stereo robot head system for accurate garment perception and manipulation. We also proposed a generic garment surface analysis approach with application to dual-arm on-table flattening. From the experimental validation, we concluded that: firstly, our stereo robot head outperforms kinect-like depth sensors in garment sensing tasks, specially for precise manipulation; secondly, the proposed generic surface features are able to parse the high-dimensional configuration space of garments by detecting and quantifying wrinkles; finally, the proposed dual-arm flattening strategy greatly improved garment manipulation efficiency with respect to the single-arm strategy. Even though we have used a towel to validate the methods and materials presented in this paper, our stereo vision system, the generic garment surface analysis and the dual-arm strategy are generic enough to be applied into different clothing (i.e. T-shirt, pants, etc.). Flattening other types of clothing usually requires more iterations than a simple towel since wrinkles are highly crumpled and disordered. For the purpose of this paper, we present a preliminary demonstration for t-shirt flattening as shown in Fig. 8 and in the video demonstration.

As future work, we intend to apply our generic garment surface analysis approach to garment grasping from a pile. Moreover, we will integrate the presented pipeline 2 into an autonomous laundry scenario (including grasping from a heap, separating, unfolding, flattening, folding, etc.). Finally, we are going to extend our proposed surface features to the topographic description of rigid and non-rigid objects for the interaction with these objects and, consequently, to understand the contents of a scene for human-robot interaction and cognitive applications.

REFERENCES

- [1] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 2308–2315.
- [2] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1703–1708.
- [3] —, "Findddd: A fast 3d descriptor to characterize textiles for robot manipulation," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 824–830.
- [4] B. Willimon, S. Birchfield, and I. Walker, "Classification of clothing using interactive perception," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1862–1868.
- [5] B. Willimon, I. Walker, and S. Birchfield, "A new approach to clothing classification using mid-level layers," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 4271–4278.
- [6] Y. Li, C.-F. Chen, and P. K. Allen, "Recognition of deformable object category and pose," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2014.
- [7] Y. Li, Y. Wang, M. Case, S.-F. Chang, and P. K. Allen, "Real-time pose estimation of deformable objects using a volumetric approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 1046–1052.
- [8] P. Monso, G. Alenya, and C. Torras, "Pomdp approach to robotized clothes separation," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 1324–1329.
- [9] J. Van Den Berg, S. Miller, K. Goldberg, and P. Abbeel, "Gravity-based robotic cloth folding," in *Algorithmic Foundations of Robotics IX*. Springer, 2011, pp. 409–424.
- [10] S. Miller, J. Van Den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 249–267, 2012.
- [11] J. Stria, D. Průša, V. Hlaváč, L. Wagner, V. Petrík, P. Krsek, and V. Smutný, "Garment perception and its folding using a dual-arm robot," in *Proc. International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 9 2014, pp. 61–67.
- [12] M. Cusumano-Towner, A. Singh, S. Miller, J. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 3893–3900.
- [13] B. Willimon, S. Birchfield, and I. D. Walker, "Model for unfolding laundry using interactive perception," in *IROS*, 2011, pp. 4871–4876.
- [14] A. Doumanoglou, A. Kargakos, T.-K. Kim, and S. Malassiotis, "Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 987–993.
- [15] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis, "Active random forests: An application to autonomous unfolding of clothes," in *Computer Vision ECCV 2014*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8693, pp. 644–658. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10602-1_42
- [16] J. Stria, D. Průša, and V. Hlaváč, "Polygonal models for clothing," in *Proc. Towards Autonomous Robotic System (TAROS)*, ser. Lecture Notes in Artificial Intelligence, vol. 8717. Springer, 9 2014, pp. 173–184.
- [17] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009, p. 5.
- [18] G. Aragon-Camarasa, H. Fattah, and J. P. Siebert, "Towards a unified visual framework in a binocular active robot vision system," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 276–286, Mar. 2010.
- [19] G. Aragon-Camarasa and J. P. Siebert, "Unsupervised clustering in hough space for recognition of multiple instances of the same object in a cluttered scene," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1274–1284, 2010.
- [20] P. Cockshott, S. Oehler, T. Xu, P. Siebert, and G. Aragon-Camarasa, "A parallel stereo vision algorithm," in *Many-Core Applications Research Community Symposium 2012*, 2012.
- [21] B. F. Gregorski, B. Hamann, and K. I. Joy, "Triangulation," in *Proc. ARPA Image Understanding Workshop*, 1994, pp. 957–966.
- [22] G. Aragon-Camarasa, S. B. Oehler, Y. Liu, S. Li, P. Cockshott, and J. P. Siebert, "Glasgow's stereo image database of garments," *arXiv preprint arXiv:1311.7295*, 2013.
- [23] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and vision computing*, vol. 10, no. 8, pp. 557–564, 1992.
- [24] Y. Ohtake, A. Belyaev, and H.-P. Seidel, "Ridge-valley lines on meshes via implicit surface fitting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 609–612, Aug. 2004. [Online]. Available: <http://doi.acm.org/10.1145/1015706.1015768>
- [25] A. Belyaev and E. Anoshkina, "Detection of surface creases in range data," in *Mathematics of Surfaces XI*. Springer, 2005, pp. 50–61.
- [26] J. A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Cambridge university press, 1999, vol. 3.